Characterizing Patients with Type 2 Diabetes Mellitus Treated with Anti-Diabetic Medication: A Feasibility Study to Enable Future Multi-Database Application

<Jiwon Um[1,2]>, <Seng Chan You[1,2] >

[1]Department of Biomedical Systems Informatics, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Republic of Korea

Background

Type 2 diabetes mellitus (T2DM) represents a global health crisis affecting over 537 million people worldwide, with Asia-Pacific regions bearing a disproportionate burden. China and India account for approximately 40% of all people with diabetes globally [1], yet diabetes affects younger populations in Asia (40-59 years) compared to Europe (60+ years).

Despite this overwhelming disease burden, the evidence based for T2DM treatment remains predominantly derived from Western population. Analysis of 405 large diabetes RCTs revealed only 24.0% enrollment of non-White populations [2], while South Asians contribute merely 2.1% to global diabetes research output despite representing a substantial proportion of the global diabetic population [3]. The landmark LEGEND-T2DM study encompassed > 190 million patients in the US but only ~50 million internationally [4], highlighted Western-centric evidence generation.

The 2025 OHDSI APAC studies encompass three distinct research questions:

- Association Between Fasting Plasma Glucose Levels and Annual Hospitalization Days: A Multicenter Study Using the OHDSI Framework
- Safety Outcomes of Semaglutide in Type 2 Diabetes Using Regional Health Data: A Target Trial Emulation
- Studying the Disease Trajectory of Type 2 Diabetes with Transformer-based Model

These studies represent the OHDSI network research driven by OHDSI-China. We aimed to conduct feasibility assessment for 2025 OHDSI APAC studies focusing on type 2 diabetes.

Methods

We conducted a feasibility analysis using the OMOP CDM version 5.4 of the Yonsei University Health System (YUHS) database, containing longitudinal health records of 6,401,087 patients (January 2006 to December 2023). The study population consisted of patients with T2DM identified based on documented diagnoses or diabetes-related complications not attributable to other subtypes. This feasibility study focused on cohort characterization without specific treatment exposure definitions, as the primary objective was to assess data availability and quality for future comparative effectiveness research. Descriptive cohort characteristics included distributions of age, gender, comorbidities, laboratory test results associated with T2DM and prescribed medications. These features, aligned with the objectives of the three research questions posed in the 2025 OHDSI APAC studies, were incorporated into the cohort characterization process.

CohortDiagnostics [5], an R package developed by the OHDSI community, was used to evaluate the validity, consistency, and feasibility of the T2DM cohort phenotype. The tool provides standardized diagnostics for cohort characterization, misclassification detection, and cross-database comparison. All cohort definitions and codes used in this study are publicly available.

Cohort characteristics were analyzed using standard descriptive statistics, with frequencies and percentages calculated for categorical variables and means and standard deviations for continuous measurements over a 365-day baseline period.

Results

CohortDiagnostics was performed to assess the feasibility of T2DM phenotypes in the YUHS database. A total of 115,935 T2DM patients was identified and characterized. The baseline clinical characteristics of study population included demographics, comorbidities, and lab measurements.

Key laboratory parameters yielded clinically appropriate ranges across multiple sample types: serum/plasma glucose (145.9±65.7 mg/dL), post-prandial glucose (206.4±81.7 mg/dL), HbA1c (7.2±1.4%), and triglycerides (149.7±123.3 mg/dL). Drug utilization patterns showed metformin as the most prescribed medication (19.45%), followed by insulin (11.76%) and DPP-4 inhibitors (10.89%).

Table 1. Baseline clinical characteristics of T2DM cohort

Characteristics	*N* = 115935
Gender[a]	
Female	50201 (43.30)
Male	65734 (56.70)
Age[a], yr	
0-19	678 (0.58)
20-29	1300 (1.12)
30-39	1300 (3.64)
40-49	4221 (10.06)
50-59	17860 (21.99)
60-69	32128 (30.66)
70-79	33912 (23.91)
80-89	18192 (7.51)
90-99	2420 (0.53)
Comorbidity[a]	
Acute myocardial infarction	2626 (2.27)
Acute pancreatitis	723 (0.62)
Chronic kidney disease due to T2DM	302 (0.26)
Heart failure	7052 (6.08)
Ischemic stroke	51 (0.04)
Peripheral vascular disease	4935 (4.26)
Diabetic retinopathy	8855 (7.64)
Drug Use[a]	
Insulin	40758 (11.76)
Metformin	58703 (19.45)

Alpha glucosidase inhibitors	2069 (0.69)
Dipeptidyl peptidase 4 (DPP-4) inhibitors	32869 (10.89)
Glucagon-like peptide-1 (GLP-1) analogues	1231 (0.41)
Sodium-glucose co-transporter (SGLT2) inhibitors	8854 (2.93)
Sulfonylureas	22389 (7.42)
Thiazolidinediones	4992 (1.65)
Combinations of oral blood glucose lowering drugs	71172 (23.59)
Blood Glucose[b]	
Glucose mean value in blood estimated from glycated hemoglobin	159.9±42.4
Glucose in serum or plasma	145.9±65.7
Glucose in serum or plasma-2 hours after meal	206.4±81.7
Glucose in arterial blood	126.9±43.7
Glucose in capillary blood by glucometer	144.7±61.0
HbA1c[b] (%)	7.2±1.4
Triglyceride[b]	
Triglyceride in pleural fluid	20.7±192.1
Triglyceride in serum or plasma	149.7±123.3
Triglyceride in serum or plasma- 12-hour fasting	135.7±85.5

[a] Values are presented as number (percentage)

[b] Values are presented as mean ± standard deviation

Conclusion

Preliminary feasibility assessment of patients with T2DM using the YUHS database (*N*=115,935) demonstrated successful mapping of key clinical features including demographics, comorbidities, and laboratory measurements essential for T2DM research. While some measurement timing precision issues were identified, the overall data completeness supports feasibility of the planned APAC studies. This validation establishes methodological foundations for deploying OHDSI cohort definitions across Asia-Pacific CDM environments in the 2025 collaborative diabetes studies. Following this validation, the pipeline will be deployed across three Chinese CDM databases: The Ningbo Regional Health Information Platform (NRHIP), the China Renal Data System (CRDS), and the Zhongshan Hospital dataset.

References

1. Gardner DSL, Saboo B, Kesavadev J, et al. Digital Health Technology in Diabetes Management in the Asia–Pacific Region: A Narrative Review of the Current Scenario and Future Outlook. Diabetes Ther. 2025;16:349-369. doi:10.1007/s13300-025-01692-0

2. Zhang J, Van Spall HGC, Wang Y, et al. Twenty-year trends in racial and ethnic enrollment in large diabetes randomized controlled trials. BMC Med. 2022;20(1):294. doi:10.1186/s12916-022-02501-2

3. Ranasinghe P, Jayawardena R, Katulanda P. Diabetes mellitus in South Asia: scientific evaluation of the research output. J Diabetes. 2013;5(1):34-42. doi:10.1111/1753-0407.12003

4. Suchard MA, Schuemie MJ, Krumholz HM, et al. Large-scale evidence generation and evaluation across a network of databases for type 2 diabetes mellitus (LEGEND-T2DM): a protocol for a series of multinational, real-world comparative cardiovascular effectiveness and safety studies. BMJ Open. 2022;12(6):e057977. doi:10.1136/bmjopen-2021-057977

5. Schuemie MJ, Cepeda MS, Suchard MA, et al. CohortDiagnostics: Assessing Cohort Construction and Evaluation of Phenotype Algorithms. OHDSI Collaborative. Available from: https://ohdsi.github.io/CohortDiagnostics/